DOCUMENT RESUME

ED 201 664 TM 810 257

AUTHOR Dunn-Rankin, Peter; And Others
TITLE Interjudge Differences Following Free Clustering.
PUB DATE [81]
NOTE 30p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cluster Analysis; *Cluster Grouping; *Computer Oriented Programs; Mathematical Formulas; Multidimensional Scaling; Probability
IDENTIFIERS *Inter Rater Reliability

ABSTRACT

Measuring object similarity using the method of free clustering is gaining in popularity. Instructions are usually simple and since no structure is imposed on the subject's selection, response bias is reduced. More importantly, measures of object similarity derived from the judges' clustering can be adequately analyzed by the methods of multidimensional scaling. Subjects are usually asked to judge or select objects which are most similar and form them into groups or clusters. The number of clusters is variable and an object by itself can constitute a cluster. This paper is concerned with measuring and displaying distributions of differences between judges following free clustering. By using a computer to generate random distributions of differences, the probability distributions for various numbers of objects in a cluster were determined and are presented. Theoretical analyses are used to create general formulas for the mean and standard deviations of the differences for varying the number of objects. The results should have value for assessing interjudge reliability and for analyzing object similarity matrices since the responses of groups of different judges can be analyzed separately. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED201664

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

INTERJUDGE DIFFERENCES FOLLOWING FREE CLUSTERING

Peter Dunn-Rankin, Eddie Wong, and Ralph Freese
University of Hawaii at Manoa

ABSTRACT

This paper is concerned with measuring and displaying distributions of differences between judges following free clustering. By using a computer to generate random distributions of differences the probability distributions for k = 3(1) 20 objects were determined and are presented. Theoretical analyses are used to create general formulas for the mean and standard deviations of the differences for varying k. The results should have value for assessing interjudge reliability and for analyzing object similarity matrices since the responses of groups of different judges could be analyzed separately.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
P. Dunn-Rankin
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 267

INTERJUDGE DIFFERENCES FOLLOWING FREE CLUSTERING

Peter Dunn-Rankin, Eddie Wong, and Ralph Freese
University of Hawaii at Manoa

Object Similarity

Measuring object similarity using the method of free clustering is gaining in popularity. Instructions are usually simple and since no structure is imposed on the subject's selection, response bias is reduced. More importantly, measures of object similarity derived from the judges' clustering can be adequately analyzed by the methods of multidimensional scaling.

Subjects are usually asked to judge or select objects which are most similar and form them into groups or clusters. The number of clusters is variable and an object by itself can constitute a cluster. Suppose there are k objects and n judges and let a, b, c, and d denote typical objects while i and j are typical judges. We number the clusters of each judge; i(a) denotes the number of the cluster into which judge i places object a. A measure of proximity or similarity between any two objects is found by calculating their "percent overlap", that is, the number of times the judges have placed the objects in the same group divided by the number of judges. More formally,

$$\text{if } i(a) = i(b), \text{ then } x_i^{(ab)} = 1$$

$$\text{if } i(a) \neq i(b), \text{ then } x_i^{(ab)} = 0$$

$$S_{ab} = \sum_{i=1}^{n} x_i^{(ab)}/n.$$

Interjudge Differences

The differences between judges' free clustering behavior, however, has not been fully explored. Often, judges will not cluster objects in similar ways.

It is a question of interest whether some judges should be excluded or separated from the rest and the analysis done over separate sets of judges.

One way to determine the difference between two judges is to pair the objects in all possible $\binom{k}{2}$ ways and create a vector or matrix of ones and zeros for each judge by providing a score of one when paired group numbers are the same and zero otherwise (the same group numbers occur when the objects have been placed in the same group). The difference between two judges is simply the sum of the squared or absolute differences between these matrices (Boorman and Arabie, 1972). More formally X is a function of two judges, i and j, defined by

$$d_{ij} = X(i,j) = |\{(a,b): 1 \leq a < b \leq k \text{ and } i(a) = i(b) \text{ and } j(a) \neq j(b) \text{ or } i)a) \neq i(b) \text{ and } j(a) = j(b)\}|$$

Notice X(i,j) simply counts the number of pairs of objects which one judge thought were similar and the other did not. If we let

$$(1) \quad X_{ab}(i,j) = \begin{cases} 1 & \text{if } i(a) = i(b) \text{ and } j(a) \neq j(b) \\ 1 & \text{if } i(a) \neq i(b) \text{ and } j(a) = j(b) \\ 0 & \text{otherwise} \end{cases}$$

then

$$(2) \quad d_{ij} = X(i,j) = \sum_{1 \leq a < b \leq k} X_{ab}(i,j).$$

Distributions of Random Differences

Suppose three objects are to be clustered. Any time a judge groups two or more objects together the members of the cluster become indistinguishable. The chance probability of the difference between two judges can be determined by assuming a random uniform distribution of letters, each representing an object. If two or more identical letters are randomly drawn they constitute a random

cluster. For three objects (k=3) the letters A, B and C represent the objects. If one assumes random selection with replacement, k^k or 27 different arrangements of the letters can occur. Table 1 presents the complete enumeration of possible differences between two judges for k = 3 objects.

TABLE 1

Possible Distribution of Differences in Free Clustering
when k = 3

Possible Clusters	AAA	BBB	CCC	AAB	AAC	BBA	BBC	CCA	CCB	ABA	ACA	BAB	BCB	CAC	CBC	BAA	CAA	ABB	CAB	ACC	BCC	ABC	ACB	BAC	BCA	CAB	CBA
AAA	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
BBB	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
CCC	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
AAB	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
AAC	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
BBA	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
BBC	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
CCA	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
CCB	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1
ABA	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	1	1	1	1	1	1
ACA	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	1	1	1	1	1	1
BAB	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	1	1	1	1	1	1
BCB	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	1	1	1	1	1	1
CAC	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	1	1	1	1	1	1
CBC	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	1	1	1	1	1	1
BAA	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1
CAA	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1
ABB	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1
CBB	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1
ACC	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1
BCC	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	1	1	1	1	1	1
ABC	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
ACB	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
BAC	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
BCA	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
CAB	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
CBA	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0

Since k = 3 there are three possible letters (A, B and C) that can be selected randomly. If one draws an A on each of three consecutive trials, all objects form one cluster. This is also true if a B is drawn in three consecutive trials. The partition (A,A,A) generates the same random clustering as (B,B,B). There are, therefore, only five possible groupings of three objects: (A,A,A); (A,A,B); (A,B,A); (B,A,A); and (A,B,C). The sum of differences between each pair of different arrangements can be 0,1,2, or 3.

Table 2 presents the cumulative probability distributions derived from the enumeration of differences in Table 1. Table 2 suggests that if one judge places three objects all together in one group and another groups each element or object separately, the difference between these two clusterings would not occur very often by chance. That is, it would only occur with a probability of .049.

TABLE 2

Probability of Cluster Differences
when k = 3

d	f	p	cp
0	153	.210	.999
1	216	.296	.789
2	324	.444	.493
3	36	.049	.049
Σ	729		

$\mu = 1.333$ $\sigma^2 = .740$

Since most objects to be clustered are presented randomly, the order in which the objects are selected to be placed together is not considered important. Thus, if there are five objects, A, B, C, D, E, and one judge chooses to place B and C together first and another judge puts B and C together last, the two

clusterings are not considered to be different. In this initial study an object is restricted to be a member of only one group. That is, an object does not share group membership.

The probability of possible clusters given k objects can be theoretically derived. The probability of any specific combination is equal to $(1/k)^k$. This specific probability is multiplied by the number of ways the particular combination can occur among the k objects as well as the number of arrangements of the particular combinations of groups. Table 3 shows the possible distributions of clusters when k is equal to five as well as the possibility of each combination of clusters. For example: The probability of any specific combination such as AABBA = $(1/5)^5 = .2^5 = .00032$. This is multiplied by the number of ways this can be converted to a 3 + 2 combination. This can occur as EECCE, for example. It can be shown that this particular combination can arise in 20 different ways. Thus its specific probability $20 \times .2^5 = .0064$. There are a total of 52 different cluster arrangements for k = 5.

TABLE 3

Possible Distributions of
Clusters When k = 5

1	AAAAA	5 1 Arr.	.00032 x 5 ways = .0016
	AAAAB		
	AAABA		
	AABAA	4 + 1	.00032 x 20 ways = .0064 x 5
	ABAAA		arrangements = .0320
6	BAAAA	5 Arr.	
	AAABB		
	AABAB		
	*AABBA	3 + 2	.00032 x 20 ways = .0064 x 10
	ABAAB		arrangements = .0640
	ABABA		
	ABBAA		
	BAAAB		
	BAABA		
	BABAA		
16	BBAAA	10 Arr.	
17	AAABC		
	AABAC		
	AABCA	3 + 1 + 1	.00032 x 60 ways = .0192 x 10
	ABAAC		arrangements = .1920
	ABACA		
	ABCAA		
	BAAAC		
	BAACA		
	BACAA		
26	BCAAA	10 Arr.	
	AABBC		
	AABCB		
	AACBB	2 + 2 + 1	.00032 x 60 ways = .0192 x 15
	ABABC		arrangements = .2880
	ABACB		
	ABBAC		
	ABBCA		
	ABCAB		
	ABCBA		
	ACABB		
	ACBAB		
	ACBBA		
	CAABB		
	CABAB		
41	CABBA	15 Arr.	

TABLE 3, cont.

	AABCD		
	ABACD		
	ABCAD	2 + 1 + 1	.00032 x 120 ways = .0384 x 10
	ABCDA	+ 1	arrangements = .3840
	BAACD		
	BACAD		
	BACDA		
	BCAAD		
	BCADA		
51	BCDAA	10 Arr.	

52	ABCDE	1 + 1 + 1 + 1 + 1 + 0	.00032 x 120 ways = .0384
		1 Arr.	

While the probability of a particular combination can be theoretically determined the probability distribution of the differences between any two cluster combinations is more difficult. This is because a combination such as AABBC is the same as BBAAC under free clustering but is not under rules for permutations. Thus the differences are not isomorphic to the distributions of similar elements in a specific clustering. Determing the exact probability distributions of interjudge differences does not seem feasible. For increasing k, the number of possible pairs of differences increases dramatically and Monte Carlo methods were used to calculate distributions of random differences for k objects from 3(1) 20, 25, and 30.

Table 4 presents these Monte Carlo distributions in cumulative probability form. These distributions were generated using the uniform random digit generation of the IMSL Library (1980).

TABLE 4

Cumulative probability distributions (cp) of interjudge differences (d) for k = 3(1) 20, 25, 30. Distributions based on 20,000 random differences are starred (*) in the table. All other values are based on 10,000 random differences.

$$\mu = k - 2 + \frac{1}{k} \text{ and } \sigma^2 = k - 4 + \frac{7}{k} - \frac{6}{k^2} + \frac{2}{k^3}$$

k = 3

d	cp
0	1.0000
1	.7813
2	.4839
3	.0537

k = 4

d	cp
0	1.0000
1	.9234
2	.7648
3	.3672
4	.1689
5	.0216
6	.0034

k = 5*

d	cp
0	1.0000
1	.9730
2	.8984
3	.6704
4	.3865
5	.1780
6	.0676
7	.0207
8	.0019
9	.0015
10	.0002

k = 6

d	cp
0	1.0000
1	.9918
2	.9593
3	.8485
4	.6287
5	.3933
6	.2090
7	.0936
8	.0418
9	.0146
10	.0058
11	.0021
12	.0003
13	.0003
14	.0001

k = 7

d	cp
0	1.0000
1	.9967
2	.9847
3	.9327
4	.8055
5	.5986
6	.3975
7	.2182
8	.1217
9	.0536
10	.0249
11	.0092
12	.0038
13	.0013
14	.0006
15	.0001

k = 8

d	cp
0	1.0000
1	.9995
2	.9939
3	.9713
4	.9034
5	.7685
6	.5793
7	.3890
8	.2365
9	.1315
10	.0681
11	.0350
12	.0172
13	.0071
14	.0026
15	.0012
16	.0005
17	.0003
18	.0001

Table 4 - Continued

k = 9		k = 10*		k = 11	
d	cp	d	cp	d	cp
0	1.0000	0	1.0000	1	1.0000
1	.9997	1	.9998	2	.9998
2	.9979	2	.9994	3	.9979
3	.9900	3	.9964	4	.9926
4	.9550	4	.9834	5	.9741
5	.8781	5	.9432	6	.9240
6	.7422	6	.8584	7	.8329
7	.5766	7	.7194	8	.7047
8	.3986	8	.5576	9	.5478
9	.2549	9	.3994	10	.3943
10	.1568	10	.2638	11	.2662
11	.0854	11	.1618	12	.1715
12	.0458	12	.0996	13	.1046
13	.0217	13	.0541	14	.0617
14	.0123	14	.0300	15	.0356
15	.0069	15	.0148	16	.0201
16	.0034	16	.0083	17	.0118
17	.0011	17	.0042	18	.0063
18	.0004	18	.0020	19	.0038
19	.0003	19	.0012	20	.0020
20	.0003	20	.0007	21	.0011
21	.0001	21	.0003	22	.0009
22	.0000	22	.0002	23	.0005
		23	.0000	24	.0005
				25	.0003
				26	.0001
				27	.0001
				28	.0001

Table 4 - Continued

k = 12		k = 13		k = 14	
d	cp	d	cp	d	cp
1	1.0000	1	1.0000	3	1.0000
2	.9999	2	.9999	4	.9999
3	.9994	3	.9999	5	.9982
4	.9967	4	.9993	6	.9916
5	.9878	5	.9951	7	.9757
6	.9606	6	.9857	8	.9415
7	.9037	7	.9577	9	.8796
8	.8119	8	.8950	10	.7857
9	.6831	9	.8007	11	.6643
10	.5437	10	.6743	12	.5382
11	.4001	11	.5404	13	.4058
12	.2835	12	.4122	14	.2972
13	.1931	13	.2955	15	.2071
14	.1218	14	.2003	16	.1375
15	.0757	15	.1335	17	.0910
16	.0473	16	.0852	18	.0584
17	.0286	17	.0544	19	.0355
18	.0154	18	.0339	20	.0230
19	.0085	19	.0209	21	.0131
20	.0042	20	.0125	22	.0079
21	.0023	21	.0076	23	.0051
22	.0015	22	.0051	24	.0029
23	.0009	23	.0029	25	.0015
24	.0005	24	.0016	26	.0005
25	.0004	25	.0010	27	.0004
26	.0002	26	.0005	28	.0001
		27	.0001		
		28	.0001		
		29	.0001		

Table 4 - Continued

k = 15*		k = 16		k = 17	
d	cp	d	cp	d	cp
3	1.0000	4	1.0000	5	1.0000
4	.9999	5	.9999	6	.9995
5	.9994	6	.9989	7	.9985
6	.9975	7	.9965	8	.9943
7	.9913	8	.9860	9	.9827
8	.9716	9	.9629	10	.9542
9	.0330	10	.9209	11	.9072
10	.8669	11	.8524	12	.8381
11	.7726	12	.7615	13	.7440
12	.6584	13	.6481	14	.6323
13	.5320	14	.5274	15	.5184
14	.4100	15	.4171	16	.4089
15	.3024	16	.3112	17	.3122
16	.2148	17	.2296	18	.2279
17	.1491	18	.1568	19	.1647
18	.1008	19	.1052	20	.1138
19	.0700	20	.0724	21	.0758
20	.0418	21	.0470	22	.0525
21	.0259	22	.0289	23	.0331
22	.0160	23	.0188	24	.0223
23	.0098	24	.0111	25	.0138
24	.0063	25	.0060	26	.0090
25	.0034	26	.0037	27	.0059
26	.0022	27	.0022	28	.0034
27	.0014	28	.0016	29	.0021
28	.0009	29	.0007	30	.0011
29	.0005	30	.0004	31	.0004
30	.0002	31	.0002	32	.0001
31	.0001	32	.0001		
32	.0001	33	.0001		
		34	.0001		

Table 4 - Continued

k = 18		k = 19*		k = 20*	
d	cp	d	cp	d	cp
4	1.0000	6	1.0000	7	1.0000
5	.9999	7	.9998	8	.9996
6	.9999	8	.9989	9	.9982
7	.0002	9	.9964	10	.9936
8	.9971	10	.9890	11	.9837
9	.9913	11	.9726	12	.9652
10	.9783	12	.9406	13	.9326
11	.9504	13	.8912	14	.8796
12	.9020	14	.8205	15	.8085
13	.8316	15	.7294	16	.7228
14	.7397	16	.6308	17	.6214
15	.6386	17	.5220	18	.5208
16	.5284	18	.4177	19	.4217
17	.4206	19	.3260	20	.3316
18	.3225	20	.2460	21	.2550
19	.2405	21	.1786	22	.1884
20	.1710	22	.1284	23	.1342
21	.1201	23	.0884	24	.0963
22	.0817	24	.0570	25	.0665
23	.0543	25	.0378	26	.0458
24	.0357	26	.0254	27	.0301
25	.0235	27	.0170	28	.0194
26	.0147	28	.0112	29	.0126
27	.0091	29	.0076	30	.0086
28	.0062	30	.0048	31	.0060
29	.0036	31	.0030	32	.0036
30	.0026	32	.0016	33	.0022
31	.0017	33	.0011	34	.0012
32	.0012	34	.0008	35	.0009
33	.0008	35	.0004	36	.0006
34	.0006	36	.0002	37	.0002
35	.0003	37	.0002	38	.0001
36	.0001	38	.0001	39	.0001
37	.0001			40	.0001

Table 4 - Continued

k = 25		k = 30*	
d	cp	d	cp
10	1.0000	12	1.0000
11	.9999	13	.9999
12	.9993	14	.9998
13	.9979	15	.9994
14	.9934	16	.9986
15	.9851	17	.9962
16	.9699	18	.9918
17	.9405	19	.9836
18	.9006	20	.9684
19	.8446	21	.9443
20	.7745	22	.9104
21	.6995	23	.8686
22	.6085	24	.8113
23	.5177	25	.7426
24	.4268	26	.6691
25	.3458	27	.5898
26	.2774	28	.5088
27	.2136	29	.4300
28	.1632	30	.3572
29	.1180	31	.2916
30	.0863	32	.2302
31	.0617	33	.1792
32	.0450	34	.1390
33	.0307	35	.1052
34	.0222	36	.0788
35	.0144	37	.0576
36	.0099	38	.0416
37	.0072	39	.0297
38	.0049	40	.0210
39	.0029	41	.0146
40	.0020	42	.0108
41	.0012	43	.0070
42	.0008	44	.0045
43	.0006	45	.0029
44	.0004	46	.0020
45	.0002	47	.0013
46	.0001	48	.0010
		49	.0007
		50	.0004
		51	.0003
		52	.0002
		53	.0002
		54	.0001
		55	.0001

The Mean and Variances of Interjudge Differences

The theoretical mean and variance for the distributions of interjudge differences can be derived as follows:

The specific difference between two judges is $d_{ij} = \Sigma X_{ab}$ where $X_{ab}(i,j)$ is defined as in (1). $E(X_{ab})$ for a single pair (a,b) of the k objects is

$$E(X_{ab}) = 2\,\frac{1}{k}\,(1 - \frac{1}{k}) = \frac{2(k-1)}{k^2}$$

since 1/k is the probability that judge i puts a and b together and 1-(1/k) is the probability that judge j does not put a and b together. The constant 2 results because the judges' roles could be reversed.

Since X is the sum of $\binom{k}{2}$ X_{ab}'s

$$E(X) = \binom{k}{2}\,\frac{2(k-1)}{k^2} = k - 2 + \frac{1}{k}$$

Since the covariance of X_{ab} and X_{cd} is 0 (if $(a,b) \neq (c,d)$) the variance of X is given by (recall $V(X_{ab}) = E(X_{ab}^2) - E(X_{ab})^2$ and $X_{ab}^2 = X_{ab}$):

$$V(X) = \Sigma V(X_{ab})$$

$$= \binom{k}{2} 2\frac{(k-1)(k^2 - 2k + 2)}{k^4}$$

$$= \frac{k(k-1)^2(k^2 - 2k + 2)}{k^4}$$

$$= \frac{k^5 - 4k^4 + 7k^3 - 6k^2 + 2k}{k^4}$$

$$= k - 4 + \frac{7}{k} - \frac{6}{k^2} + \frac{2}{k^3}$$

These derivations for μ and σ^2 are so close to the means and variances of the sample distributions that one can conclude that the sample distributions are

accurate reflections of the cumulative probabilities of true chance differences (for k = 15, μ = 14.0625 and $\bar{X}$ = 14.068, for example).

Comparison with the Normal Distribution

If the distributions of differences were normally distributed then the theoretical mean and variance could be used with any suitable normal probability table to test for significant differences. Figure 1 illustrates the distribution of 20,000 chance differences when k = 20 objects and 30 objects, respectively. The upper tail distribution for 30 objects is quite close to the normal curve. It is suspected that as k increases the distributions will become increasingly normal. For k less than 30 the sample distributions will require larger differences to be declared significant at standard alpha levels than the normal distribution of differences using the true mean and variance.

Calculations of the sample skewness and kurtosis of interjudge differences for k = 5, 10, 15, 20, 25, and 30 objects suggest that as k increases these moments tend toward their normal values. It is also possible to theoretically derive these moments. Further study is needed to definitely assert normality for the distribution of differences with large values of k (k>30).

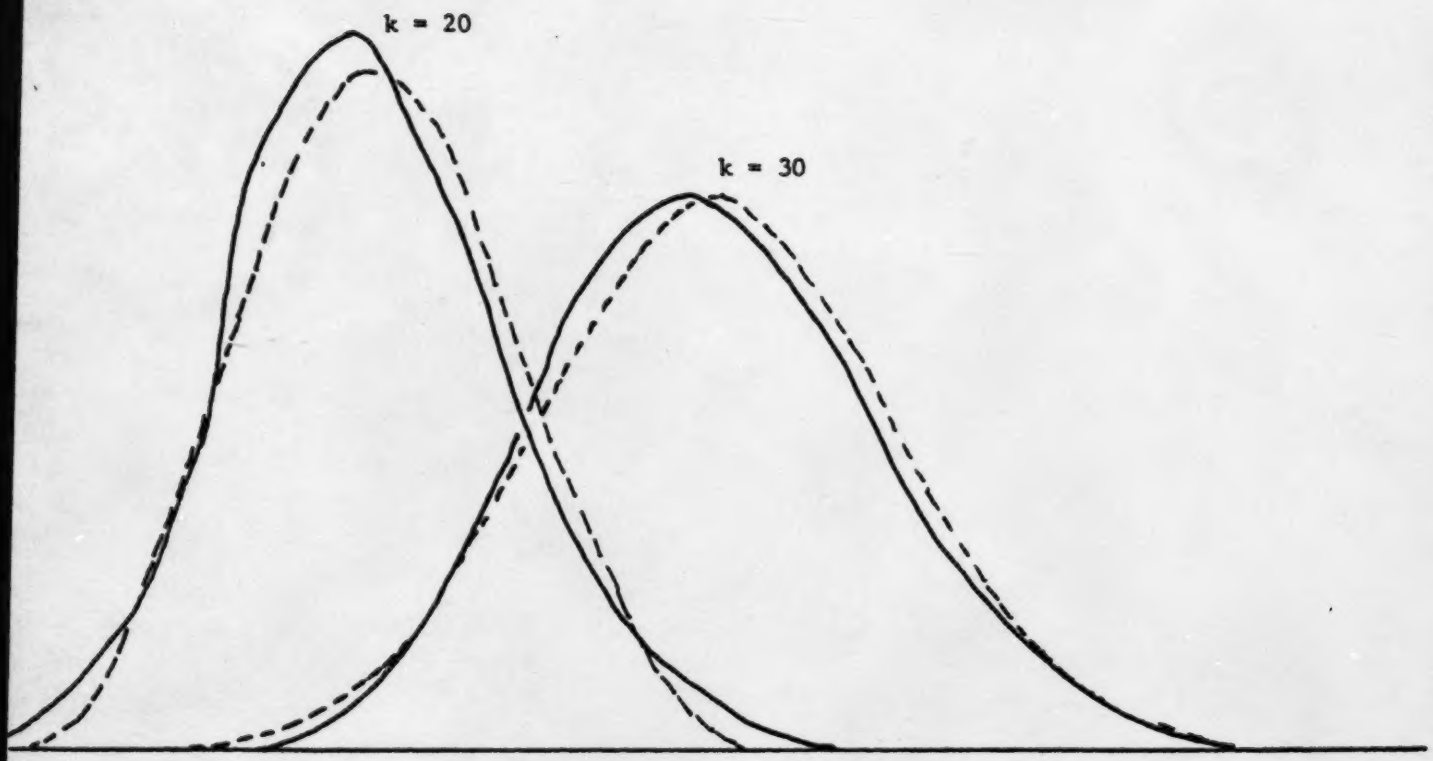

Fig. 1. The distribution of 20,000 chance differences when k = 20 and 30 objects, respectively. Dotted curve represents the normal distribution of the same data.

Example

How can the derived information be applied? Once a clustering experiment is performed how may the judges be compared? An example may be illustrative.

Fifteen graduate students clustered 11 words that began with the letter a. The words (a, as, at, and, aged, army, away, areas, almost, admits, aiming) were typed on IBM prestige elite type on small slips of paper and presented randomly. Directions were "Put the words you think are similar together in groups. You may form as many or as few groups as you wish." Table 6 provides the basic data. The entries are cluster numbers. The judges a to o are listed vertically and the words represent the columns.

TABLE 6

Group membership identification numbers for 11 words beginning with the letter a. The 15 adult respondents are listed by letter from a to o in the left-hand margin.

	a	admits	aged	almost	aiming	and	as	at	areas	army	away
a	1	2	3	2	4	3	1	1	5	6	6
b	1	2	3	2	2	1	1	1	3	2	2
c	1	2	1	2	2	3	1	1	3	3	3
d	1	2	3	4	4	1	5	6	7	8	8
e	1	2	3	4	4	1	5	6	7	8	8
f	1	2	3	3	4	5	1	6	7	8	8
g	1	2	2	2	2	3	1	1	2	2	2
h	1	2	3	4	5	6	7	7	8	9	9
i	1	2	3	2	4	5	1	6	4	4	4
j	1	2	3	4	5	2	1	1	2	6	6
k	1	2	3	2	4	1	1	1	3	5	5
l	1	2	3	4	2	3	1	1	3	3	3
m	1	2	3	2	4	5	1	1	6	7	5
n	1	2	3	2	4	5	1	1	6	7	7
o	1	2	3	2	2	3	1	1	3	2	3

Before analyzing the differences between the judges the similarity between the objects was ascertained. Percent overlap values derived by comparing the pairwise columns of objects were formed. These are presented in Table 7. The

object similarities were analyzed by multidimensional scaling (Kruskal, 1964) and a two-space configuration of the objects is shown in Figure 2. It can be seen that the objects have been classified largely on the basis of their length.

TABLE 7

Percent overlap scores for 11 words beginning with the letter a. Data is based on 15 adult subjects' free clustering.

	a	admits	aged	almost	aiming	and	as	at	areas	army	away
a	1.000	0.0	0.067	0.0	0.0	0.267	0.800	0.667	0.0	0.0	0.0
admits	0.0	1.000	0.067	0.600	0.333	0.067	0.0	0.0	0.133	0.200	0.133
aged	0.067	0.067	1.000	0.133	0.067	0.200	0.067	0.067	0.333	0.133	0.200
almost	0.0	0.600	0.133	1.000	0.400	0.0	0.0	0.0	0.067	0.200	0.133
aiming	0.0	0.333	0.067	0.400	1.000	0.0	0.0	0.0	0.133	0.267	0.200
and	0.267	0.067	0.200	0.0	0.0	1.000	0.133	0.133	0.267	0.133	0.267
as	0.800	0.0	0.067	0.0	0.0	0.133	1.000	0.733	0.0	0.0	0.0
at	0.667	0.0	0.067	0.0	0.0	0.133	0.733	1.000	0.0	0.0	0.0
areas	0.0	0.133	0.333	0.067	0.133	0.267	0.0	0.0	1.000	0.267	0.333
army	0.0	0.200	0.133	0.200	0.267	0.133	0.0	0.0	0.267	1.000	0.867
away	0.0	0.133	0.200	0.133	0.200	0.267	0.0	0.0	0.333	0.867	1.000

Did the judges, however, generally agree in their clustering behavior? Table 8 provides the differences between all pairs of judges based on definitions given in this paper. Since there are 11 objects, reference to Table 4 shows that a difference of 15 or greater would occur less than 5 times in 100 by chance. Given this frame of reference there appear to be real differences between the judges. The difference matrix was also analyzed by multidimensional methods. This result is presented as Fig. 3. Figure 3 suggests that judge G is somewhat different from the other judges. Next to each judge is placed the results of his or her individual grouping strategy. The plane is divided by judges who group words into just a few clusters and those who place words into many clusters. A scrutiny of the "north-south" dimension suggests that judges have difficulty with the word

and. Some see it as a short word, some see it as a medium sized word, and the rest are undecided.

TABLE 8

Lower triangular matrix of differences between 15 judges following their clustering of 11 words

	Subjects													
	a	b	c	d	e	f	g	h	i	j	k	l	m	n
b	13.													
c	11.	18.												
d	7.	14.	14.											
e	7.	14.	14.	0.										
f	5.	16.	14.	4.	4.									
g	20.	13.	21.	23.	23.	21.								
h	4.	15.	13.	3.	3.	3.	22.							
i	8.	15.	13.	9.	9.	7.	16.	8.						
j	5.	16.	12.	8.	8.	6.	21.	5.	11.					
k	5.	8.	14.	8.	8.	8.	21.	7.	11.	8.				
l	10.	19.	9.	15.	15.	13.	18.	12.	14.	11.	13.			
m	3.	14.	10.	8.	8.	6.	21.	5.	9.	6.	6.	11.		
n	1.	12.	10.	6.	6.	4.	19.	3.	7.	4.	4.	11.	2.	
o	11.	12.	12.	16.	16.	16.	15.	15.	15.	14.	14.	9.	10.	12.

Discussion

Since scaling is relative, it is helpful to have an objective means of comparing the differences between judges. Intuitively, however, it seems that the null hypothesis of chance differences may make unreasonable assumptions about experiments of this type. It seems apparent that someone could put all the words together in one cluster based on ideas such as "they are all words, they all begin with a, they are all familiar, etc." Someone else, on the other hand, could defend a clustering of 11 separate cluster contending that "they all have different meanings, etc." The chance probability of this difference between two such judges (a difference of 55) is almost impossible, yet its occurrence in the real world is reasonably possible.

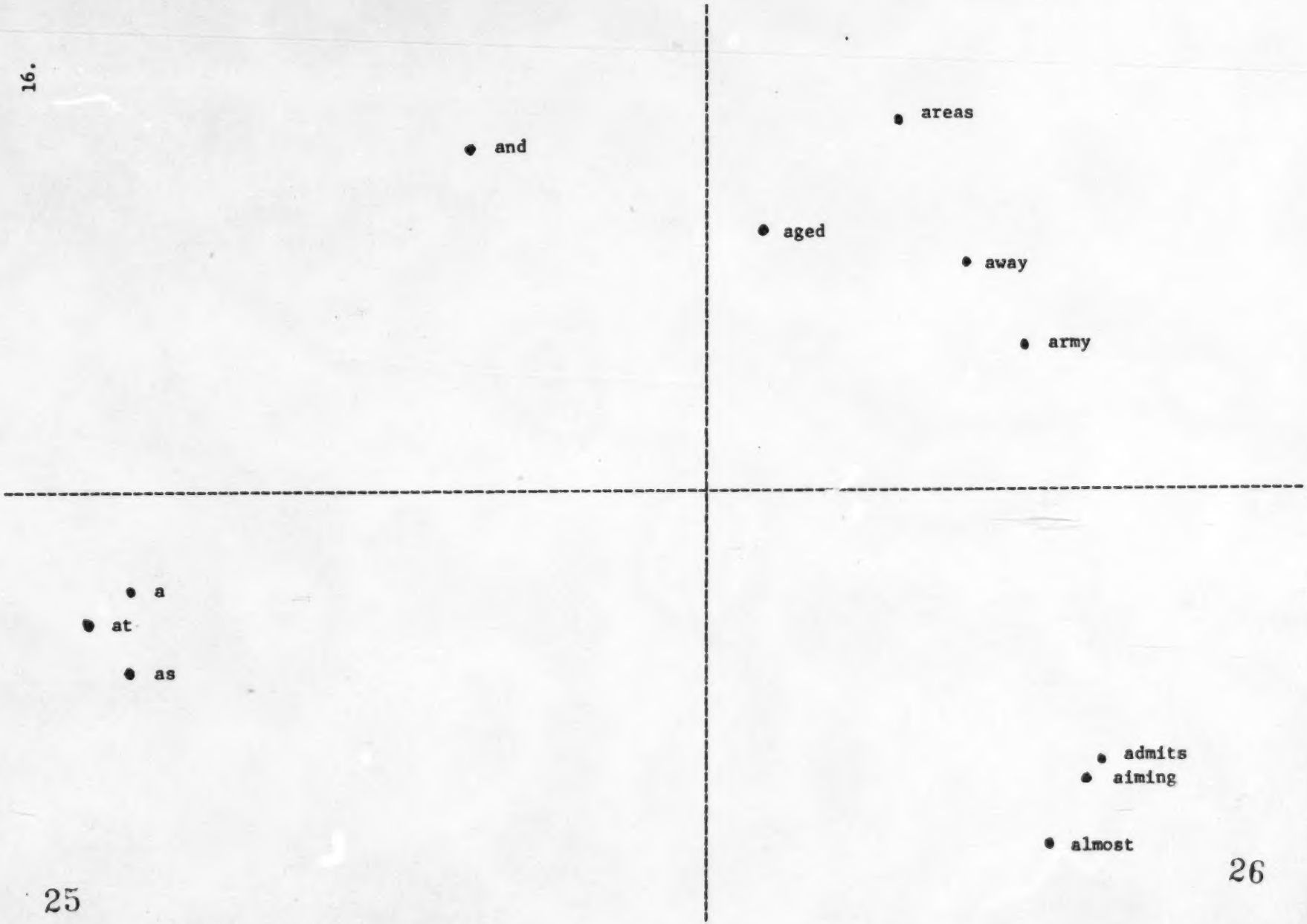

Figure 2. Multidimensional plot of 11 words beginning with a. Percent overlap data were derived from free cluster responses of 15 adults.

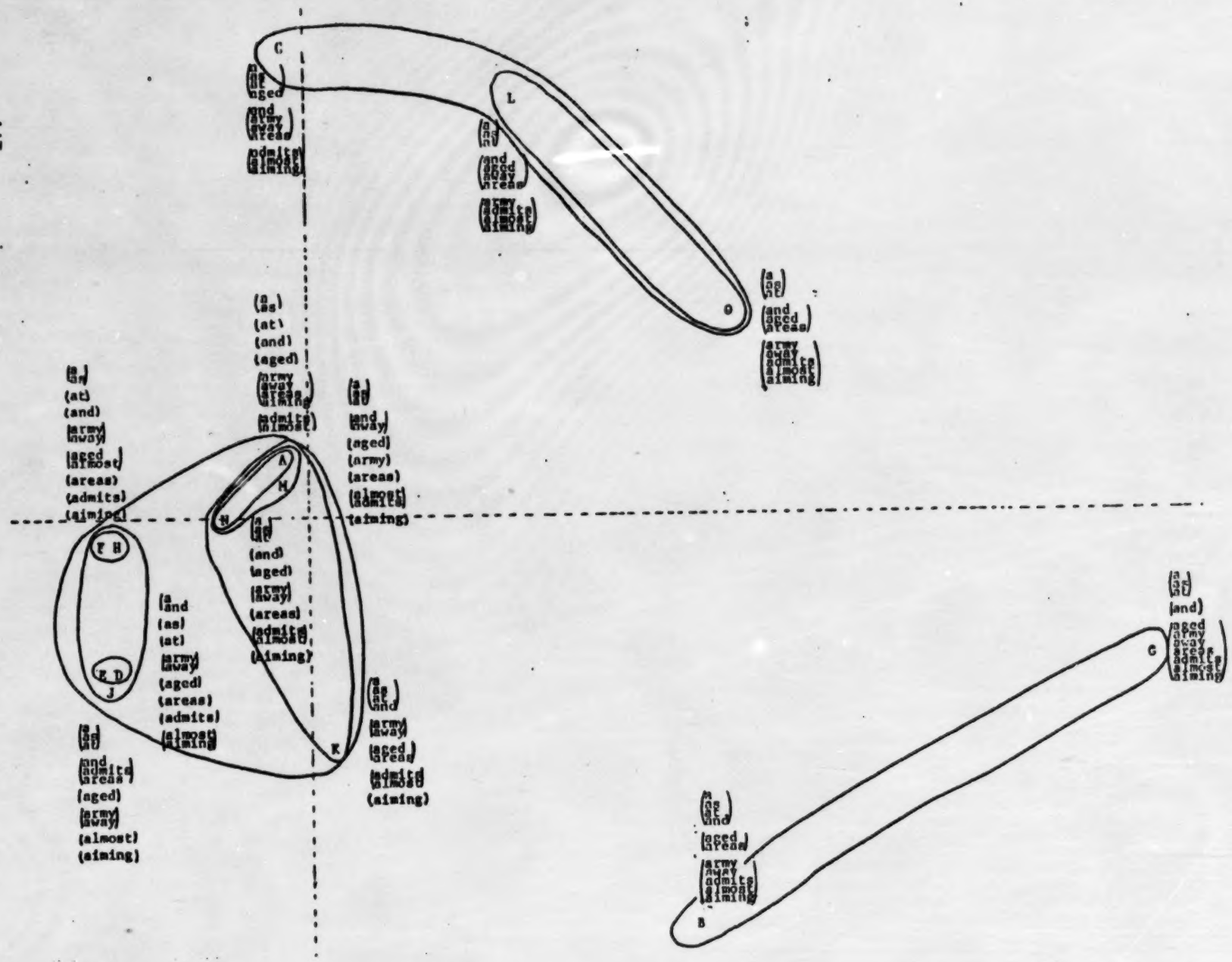

Figure 3. The nonmetric multidimensional scale analysis of differences between judges following the free clustering of 11 words. A hierarchical clustering is superimposed on the two dimensional solution.

If the distributions of differences become normal after k reaches 30 then multiple comparison methods could be applied to lower triangular matrices of differences since the true variance is known. Thus groups of judges significantly different from each other could be analyzed separately or outlying judges could be excluded from subsequent analyses of the objects. In the example above one might wish to exclude subject G or group all similar judges together and analyze the objects over these separate groups.

References

Boorman, S. A. and Arabie, P. Structural measures and the method of sorting. In R. N. Shepard, & S. B. Nerlove (Eds.), Multidimensional Scaling (Vol. 1). New York: Seminar Press, 1972.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, March 1964, 29, 1-27.

International Mathematical and Statistical Library, 8th ed., 1980. IMSL, Inc., Houston, Texas.